



A straightforward multiallelic significance test for the Hardy-Weinberg equilibrium law

Marcelo S. Lauretto¹, Fabio Nakano², Silvio R. Faria Jr³, Carlos A.B. Pereira³ and Julio M. Stern³

¹Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, Brazil.

²Instituto de Química, Universidade de São Paulo, São Paulo, Brazil.

³Instituto de Matemática e Estatística Universidade de São Paulo, São Paulo, Brazil.

Abstract

Much forensic inference based upon DNA evidence is made assuming Hardy-Weinberg Equilibrium (HWE) for the genetic loci being used. Several statistical tests to detect and measure deviation from HWE have been devised, and their limitations become more obvious when testing for deviation within multiallelic DNA loci. The most popular methods-Chi-square and Likelihood-ratio tests-are based on asymptotic results and cannot guarantee a good performance in the presence of low frequency genotypes. Since the parameter space dimension increases at a quadratic rate on the number of alleles, some authors suggest applying sequential methods, where the multiallelic case is reformulated as a sequence of "biallelic" tests. However, in this approach it is not obvious how to assess the general evidence of the original hypothesis; nor is it clear how to establish the significance level for its acceptance/rejection. In this work, we introduce a straightforward method for the multiallelic HWE test, which overcomes the aforementioned issues of sequential methods. The core theory for the proposed method is given by the Full Bayesian Significance Test (FBST), an intuitive Bayesian approach which does not assign positive probabilities to zero measure sets when testing sharp hypotheses. We compare FBST performance to Chi-square, Likelihood-ratio and Markov chain tests, in three numerical experiments. The results suggest that FBST is a robust and high performance method for the HWE test, even in the presence of several alleles and small sample sizes.

Key words: Hardy-Weinberg equilibrium, significance tests, FBST.

Received: November 4, 2008; Accepted: May 25, 2009.

Introduction

The Hardy-Weinberg law is one of most important principles in population genetics, and establishes a direct relationship between allele and genotypic proportions in a population. This law states that in a large population of panmictic dioecious organisms with non-overlapping generations, the allelic and genotypic frequencies at a locus will stay unchanged, provided that migration, mutation, and natural selection do not affect that locus. When these conditions hold, it is said that the locus is under Hardy-Weinberg Equilibrium (HWE).

This principle was discussed independently by Yule, Pearson and Castle (between 1902 and 1904), for some particular allele frequencies (see references in Crow and Kimura, 1972). In 1908, Godfrey Hardy presented the general principle for two alleles (Hardy, 1908). This principle was called Hardy's law for 35 years, until Stern (1943) called attention to an article of Weinberg (1908) showing the same

principle at the same time and demonstrating its validity for multiple alleles (Crow, 1999).

Since its postulation, several results in population genetics and much forensic inference based upon DNA evidence have been based on the assumption that HWE is valid for the genetic loci of interest. Some statistical tests to detect and measure deviation from HWE have been devised, and their limitations have become more obvious when testing for deviation within multiallelic DNA loci. The most common approach consists of goodness-of-fits tests, like Chi-square and Likelihood-ratio, which are heavily based on asymptotic results, and can sometimes lead to false rejection or acceptance of HWE when the sample sizes are small and/or some genotype sample frequencies are very small (Emigh, 1980). Another approach involves exact tests, but is restricted to small dimensions and allele numbers.

A Bayesian sequential method for multiallelic HWE test was proposed by Pereira *et al.* (2006), who suggested reformulating the multiallelic case as a sequence of "biallelic" tests. In that work, the central component is the Full Bayesian Significance Test (FBST), an intuitive Bayesian approach which does not assign positive probabilities to

Send correspondence to Marcelo S. Lauretto. Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, Rua Arlindo Béttio 1000, Ermelino Matarazzo, 03828-000 São Paulo, SP, Brazil. E-mail: marcelolauretto@usp.br.

zero measure sets when testing sharp hypotheses (Pereira and Stern, 1999). Although the sequential method avoids the quadratic increase of parameter space dimension with respect to the number of alleles, it is not obvious how to assess the general evidence of the original hypothesis; nor is it clear how to establish the significance level for its acceptance or rejection (see DeGroot, 1970).

In this work, we propose a method for the multiallelic HWE test, based on the FBST. FBST has many theoretic and practical advantages over other approaches, and it has shown to be robust in several high-dimensional problems (Lauretto et al., 2008).

Background

In this section we introduce some notations, and the Hardy-Weinberg Equilibrium (HWE) formula. Let us consider k alleles A_1, A_2, \dots, A_k in a *locus*. The main interest is to assess the population relative frequencies of the genotypes $A_i A_j$ ($i, j = 1, 2, \dots, k$) which we denote by p_{ij} . As usual in the literature (see Hardy, 1908), we assume that the allele frequencies do not depend on sex and thus are symmetric, that is, $A_i A_j$ is equivalent to $A_j A_i$ and $p_{ij} = p_{ji}$. Therefore, the parameter of interest is the (lower triangular) matrix of genotype proportions:

$$\theta = (\theta_{ij}), \text{ with } \theta_{ii} = p_{ii}, \theta_{ij} = 2p_{ij} \text{ for } 1 \leq j \leq i \leq k.$$

We denote by p_1, p_2, \dots, p_k the (unknown) population frequencies of alleles A_1, A_2, \dots, A_k , with $p_i \geq 0$ and $\sum_{i=1}^k p_i = 1$. When the locus is under HWE, the genotype proportions are as follows:

$$\theta_{ii} = p_i^2; \theta_{ij} = 2p_i p_j, \quad 1 \leq j \leq i \leq k. \quad (1)$$

In order to test the HWE in a locus, one considers a sample of n individuals drawn randomly from the population. Such a sample can be presented as the array x , whose elements x_{ij} , $1 \leq j \leq i \leq k$, are counts of genotypes $A_i A_j$. The sample size n is $n = \sum_{j=1}^k x_{ij}$, and the sample frequency of allele A_i is $n_i = \sum_{j=1}^i x_{ij} + \sum_{j=i}^k x_{ji}$. Note that $\sum_i n_i = 2n$, since all loci have two alleles and this sum is the number of alleles in the whole sample. The sample proportion of allele A_i , \tilde{p}_i , is given by

$$\tilde{p}_i = \frac{n_i}{2n}. \quad (2)$$

Assuming that each individual genotype does not depend on remaining individuals in the same generation, we can consider that the genotype frequencies x_{ij} follow a multinomial distribution with unknown parameter θ ,

$$P(x|\theta) = n! \prod_{j \leq i=1}^k \frac{\theta_{ij}^{x_{ij}}}{x_{ij}!}. \quad (3)$$

Testing Procedures

In this section we present three tests used in our comparative study. These and other approaches are described by Emigh (1980), Guo and Thompson (1992), Hernández and Weir (1989) and Montoya-Delgado et al. (2001).

Chi-square goodness-of-fit test

This test involves calculating the sample chi square value,

$$\chi^2 = \sum_{j \leq i=1}^k \frac{(x_{ij} - E_{ij})^2}{E_{ij}}, \quad (4)$$

with

$$E_{ii} = np_i^2 = \frac{1}{4} \frac{n_i^2}{n} (i = 1, \dots, k),$$

$$E_{ij} = 2np_i \tilde{p}_j = \frac{1}{2} \frac{n_i n_j}{n} (j < i).$$

Under HWE, this quantity has a chi-square distribution with $k(k-1)/2$ degrees of freedom.

The Chi-square goodness-of-fit test with continuity correction involves calculating the previous statistics, with the subtraction in each term of a correction constant c :

$$\chi^2 = \sum_{j \leq i=1}^k \frac{(|x_{ij} - E_{ij}| - c)^2}{E_{ij}}. \quad (5)$$

Usually $c = 0.5$ is the value chosen.

Likelihood-ratio tests

The likelihood function, given a sample, follows directly from the multinomial distribution presented in Eq. (3). A Likelihood-ratio test is constructed by comparing the likelihood maximized under the hypothesis, L_0 , with the maximum likelihood, L_1 , not constrained by the hypothesis. For HWE we have

$$L_1 = \frac{n!}{n^n} \prod_{j \leq i=1}^k \frac{x_{ij}^{x_{ij}}}{x_{ij}!}, \quad L_0 = n! \prod_{i=1}^k \frac{\tilde{p}_i^{2x_{ii}}}{x_{ii}!} \prod_{j < i} \frac{(2\tilde{p}_i \tilde{p}_j)^{x_{ij}}}{x_{ij}!}. \quad (6)$$

with the sample allelic frequencies, \tilde{p}_i , given by Eq. (2).

The test statistic

$$G^2 = -2 \ln \left(\frac{L_0}{L_1} \right) \quad (7)$$

is asymptotically distributed as a chi-square distribution with $k(k-1)/2$ degrees of freedom.

Markov Chain Monte Carlo (MCMC) method

Proposed by Guo and Thompson (1992), the method consists of an adaptation of the Metropolis algorithm, with the construction of a Markov chain with equilibrium distribution matching the genotype probabilities under HWE of

samples that have the same allelic counts as the observed data.

Under HWE and conditional on sample allele counts, n_1, n_2, \dots, n_k , the probability of obtaining the sample x is (see Levene, 1949):

$$\Pr(x|n_1, \dots, n_k) = \frac{n! \prod_{i=1}^k n_i}{(2n)! \prod_{j < i} x_{ij}} 2^{\sum_{j < i} x_{ij}}. \quad (8)$$

Given the data x , the test evaluates

$$P = \sum_{y \in \wp} \Pr(y), \quad (9)$$

where $\wp = \{y: \Pr(y) \leq \Pr(x), y \in \Gamma_0\}$ and

$$\Gamma_0 = \Gamma(x) = \{y: y \text{ has the same allele counts as does } x\}. \quad (10)$$

The MCMC algorithm is performed in order to estimate the probability P in E. (9). Rejection or acceptance of the null hypothesis depends on whether P is smaller than a pre-specified significance level α .

Methods

The Full Bayesian Significance Test (FBST)

The Full Bayesian Significance Test (FBST) was proposed by Pereira and Stern (1999) as a coherent and intuitive Bayesian test. It assumes that the hypothesis, H , is defined as a subset defined by inequality and equality constraints:

$$H: \theta \in \Theta_H \text{ for } \Theta_H = \{\theta \in \Theta | g(\theta) \leq 0 \wedge h(\theta) = 0\}, \text{ and } \Theta \subseteq \mathbb{R}^n. \quad (11)$$

For simplicity, we often use H for Θ_H . FBST is particularly focused on precise hypotheses: *i.e.*, $\dim(\Theta_H) < \dim(\Theta)$. In this work, $f_x(\theta)$ denotes the posterior probability density function, given the observation x . Bold $\mathbf{0}$ and $\mathbf{1}$ denote vectors of appropriate dimensions.

For the HWE test, the parametric space consists of all arrays of genotype proportions

$$\Theta = \left\{ \theta = (\theta_{ij}), 1 \leq j \leq i \leq k | \theta \geq \mathbf{0} \wedge \sum_{j \leq i=1}^k \theta_{ij} = \mathbf{1} \right\}. \quad (12)$$

The space on hypothesis is

$$H \equiv \Theta_H = \left\{ \theta \in \Theta | \exists p_1, p_2, \dots, p_k : \theta_{ii} = p_i^2 (i = 1, \dots, k), \theta_{ij} = 2p_i p_j (j < i), \mathbf{0} \leq p \leq \mathbf{1}, p' \mathbf{1} = \mathbf{1} \right\}. \quad (13)$$

As previously stated, we consider that the genotype frequencies x follow a multinomial distribution, given by Eq. (3). Taking as a priori the Dirichlet distribution with parameters $(1, 1, \dots, 1)$, *i.e.*, a uniform distribution, then the a posteriori is a Dirichlet distribution with parameters $(x_{11} + 1, x_{21} + 1, x_{22} + 1, \dots, x_{kk} + 1)$ which is proportional to the likelihood function (DeGroot, 1970):

$$f_x(\theta) = \Gamma(n+k) \prod_{j \leq i=1}^k \frac{\theta_{ij}^{x_{ij}}}{\Gamma(x_{ij}+1)} \propto \prod_{j \leq i=1}^k \theta_{ij}^{x_{ij}} \quad (14)$$

The computation of the evidence measure used on the FBST is performed in two steps:

1. *The optimization step* consists of finding the maximum (supremum) of the posterior under the null hypothesis, $\theta^* = \arg \sup_H f_x(\theta)$, $f^* = f_x(\theta^*)$.

2. *The integration step* consists of integrating the posterior density over the Tangential Set, \bar{T} , where the posterior is higher than anywhere in the hypothesis, *i.e.*,

$$\begin{aligned} \bar{E}v(H) &= \Pr(\theta \in \bar{T} | x) = \int_{\bar{T}} f_x(\theta) d\theta, \text{ where} \\ \bar{T} &= \left\{ \theta \in \Theta : f_x(\theta) > f^* \right\} \end{aligned} \quad (15)$$

$\bar{E}v(H)$ is the evidence against H , and $EV(H) = 1 - \bar{E}v(H)$ is the evidence supporting (or in favor of) H . For a better understanding of this evidence measure, Figure 1 illustrates two examples in the biallelic case, showing the null and tangential sets (Θ_H and \bar{T}). Since $\theta_{21} = 1 - \theta_{11} - \theta_{22}$, the parametric space is fully defined by homozygote proportions, θ_{11} and θ_{22} . The parameter space corresponds to the area inside the triangle. Sample genotype counts for A_{11}, A_{21}, A_{22} and $\bar{E}v(H)$ are also shown in each graph. Marker '*' represents the point θ^* of maximum a posteriori density in the constrained space Θ_H , and the level curve tangent to θ^* corresponds to \bar{T} frontier. Intuitively, if the hypothesis set is in a region of "low" posterior density (as in the example 1), then \bar{T} is "heavy" and therefore $\bar{E}v(H)$ is "large" (≈ 0.91), meaning "strong" evidence against H . On the other hand, as illustrated by the example 2, if hypothesis set is in a region of "high" posterior density, then \bar{T} is a "small" set, and hence $\bar{E}v(H)$ is "small" (≈ 0.36), meaning "weak" evidence against H .

For HWE test, the point $\theta^* = \arg \sup_H f_x(\theta)$ is given as follows.

Rewriting the posterior pdf under HWE, we have:

$$f_x(\theta | H) \propto \prod_{i=1}^k p_i^{2x_{ii}} \prod_{j < i} (2p_i p_j)^{x_{ij}}. \quad (16)$$

Taking its logarithm,

$$\begin{aligned} l_x(\theta) &\equiv \log f_x(\theta | H) \propto \sum_{i=1}^k 2n_{ii} \log p_i + \\ &\quad \sum_{j < i} n_{ii} \log(2p_i p_j) = h \log 2 + \sum_{i=1}^k n_i \log p_i, \end{aligned} \quad (17)$$

where $n_i = \sum_{j=1}^i x_{ij} + \sum_{j=i+1}^k x_{ji}$, and h is the sum of sample heterozygote frequencies, $h = \sum_{j < i} x_{ii}$.

By the constraint $\sum_{i=1}^k p_i = 1$, we have:

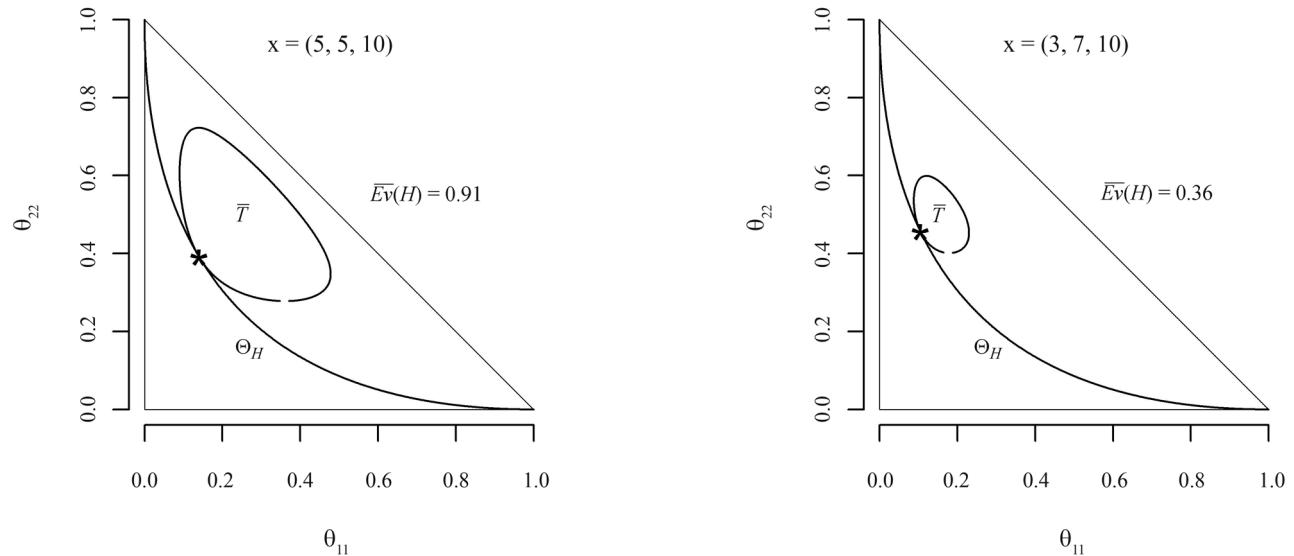


Figure 1 - Parameter space representation for two biallelic examples, $x = (5, 5, 10)$ and $x = (3, 7, 10)$, and the respective evidence against the HWE hypothesis.

$$l_x(\theta) \propto h \log 2 + \sum_{i=1}^{k-1} n_i \log p_i + \\ n_k \log \left(1 - \sum_{j=1}^{k-1} n_j \log p_j \right) \quad (18)$$

The gradients of $l_x(\theta)$ are given by:

$$\frac{\partial l_x}{\partial p_i} \propto \frac{n_i}{p_i} - \frac{n_k}{1 - \sum_{j=1}^{k-1} p_j} \quad (19)$$

Hence, the optimal point under HWE is given by the vector $p^* = (p_1^*, \dots, p_{k-1}^*)$ which satisfies:

$$Ap = b; A = \begin{pmatrix} n_1 + n_k & n_1 & \dots & n_1 \\ n_2 & n_2 + n_k & \dots & n_2 \\ \vdots & \vdots & \ddots & \vdots \\ n_{k-1} & n_{k-1} & \dots & n_{k-1} + n_k \end{pmatrix}; \quad (20)$$

$$b = \begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_{k-1} \end{pmatrix}.$$

Summing over all constraints and after some algebra, we obtain the following solution:

$$p^* = p_i^* = \tilde{p}_i = \frac{n_i}{2n}, i = 1, \dots, k.$$

The computation of θ^* from p^* follows from Eq. (1).

The integration step may be performed by generating a set of M points $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(M)}\}$ with a Dirichlet distribution with parameter $(x_{11} + 1, x_{21} + 1, x_{22} + 1, \dots, x_{kk} + 1)$

and computing the percentage of points with posterior density greater than f^* :

$$\theta^{(r)} \sim \text{Dirichlet}(x_{11} + 1, x_{21} + 1, \dots, x_{kk} + 1), r = 1, 2, \dots, M \quad (21)$$

$$\bar{Ev}(H) = \frac{\sum_{r=1}^M I(f_x(\theta^{(r)}) > f^*)}{M}$$

where $I(stat) = 1$ if $stat$ is true and 0 otherwise. A more precise and efficient Monte Carlo method for the integration step is presented by Lauretto et al. (2003).

As with any significance test, this procedure requires the choice of a threshold level, τ , for acceptance/rejection of the hypothesis at a significance level α . Several alternative methods have been developed for establishing this threshold:

- An empirical power analysis, developed by Stern and Zacks (2002) and Lauretto et al. (2003), provides critical levels that are consistent and also effective for small samples.
- A threshold based on reference sensitivity analysis and paraconsistent logic is given by Stern (2004).
- Pereira et al. (2008) relates the e-value threshold to standard p-value thresholds.
- Madruga et al. (2001) proves the existence of a loss function that renders the FBST a true Bayesian decision-theoretic procedure.
- An asymptotically consistent threshold for a given confidence level was given by Stern (2007), and Borges and Stern (2007), which we adopt in this work.

Let us consider the cumulative distribution of the evidence value against the hypothesis, $\bar{V}(\tau) = \Pr(\bar{Ev}(H) \leq \tau)$, given θ^0 , the true value of the parameter. Under appropriate

regularity conditions, for increasing sample size, we can state the following:

If H is false, $\theta^0 \notin H$, then $\bar{Ev}(H)$ converges (in probability) to one, that is, $\bar{V}(\tau) \rightarrow \delta(1)$.

If H is true, $\theta^0 \in H$, then $\bar{V}(\tau)$, the confidence level, is approximated by the function $Q(t, h, \tau) = \text{Chi2}(t - h, \text{Chi2}^{-1}(t, \tau))$, where $t = \dim(\Theta)$, $h = \dim(H)$ and $\text{Chi2}(df, x)$ denotes the cumulative chi-square distribution with df degrees of freedom.

Hence, to reject H with a significance level α , we can set $\tau = Q^{-1}(t, h, 1-\alpha)$, i.e., set τ such that $Q(t, h, \tau) = 1 - \alpha$.

Results and Discussion

The numerical experiments used in the performance analysis are based on three typical datasets. Two examples consist of simulated data used in the literature as benchmarks for comparing the performance of competing methods, while the third example is from a real dataset. These examples are presented in Figure 2, as lower triangular matrices containing genotype frequencies. The first example is taken from Louis and Dempster (1987) and consists of a sample of size 45 of genotype frequencies distributed in four alleles. The second example is given by Guo and Thompson (1992) and consists of a sample of size 30 of simulated genotype frequencies simulated under HWE with underlying allele frequencies (0.2, 0.2, 0.2, 0.2, 0.05, 0.05, 0.05, 0.05). The third example is from a rheumatoid arthritis (RA) study performed by Wordsworth *et al.* (1992), where two hundred and thirty RA patients were genotyped for the HLA-DR locus. The DR4 allele was subdivided into Dw4, Dw14 and other subtypes. DRX represents all non-DR1, non-Dw4, non-Dw14 alleles. This example is used by Chen and Thomson (1999) as a benchmark.

Our main interest is to compare the error convergence of FBST and other methods presented in this work (MCMC, Chi-square and Likelihood-ratio) for increasing

sample sizes. For each sample size $n \in \{30, 50, 100, 200\}$, we simulated two collections of 100 samples. The first collection consists of samples drawn under HWE, *i.e.*, each sample is drawn with a multinomial distribution with parameters (n, θ^*) , with $\theta^* = \arg \max_{\theta \in H} f_x(\theta)$. The second collection consists of samples drawn with a multinomial distribution with parameters $(n, \theta^{(h)})$, where $\theta^{(h)}$ is drawn under the posterior distribution. That is, each sampling iteration is performed in two steps:

- a) draw $\theta^{(h)} \sim \text{Dirichlet}(x_{11} + 1, x_{21} + 1, \dots, x_{kk} + 1)$,
 where x_{ij} ($1 \leq j \leq i \leq k$) are the frequencies in the original dataset; and
 b) draw $X^{(h)} \sim \text{Multinomial}(n, \theta^{(h)})$.

The Type I error (rejection rate of a true hypothesis) is estimated by the proportion of samples in collection 1 such that HWE is rejected, and the Type II error (acceptance rate of a false hypothesis) is estimated by the proportion of samples in collection 2 such that HWE is accepted. The performance criterion used in this work is the average error, *i.e.*, the average of Types I and II error rates. Two standard significance levels, $\alpha \in \{0.01, 0.05\}$, were used to calibrate the asymptotic acceptance/rejection threshold of each method.

A variability measure for the errors was obtained by performing 10 batches of simulations and computing the mean and standard deviation of average errors across the batches.

Figure 3 presents the average errors for FBST, MCMC, Chi-square (with continuity correction) and Likelihood ratio for simulated data based on examples 1, 2 and 3. The bar height represents the mean of average errors, and the vertical line on the top of each bar is the error bar, representing the mean \pm one standard deviation of average errors.

For simulated data based on examples 1 and 3, the best competitors are FBST and Likelihood-ratio test, while for simulated data based on example 2, the best competitors

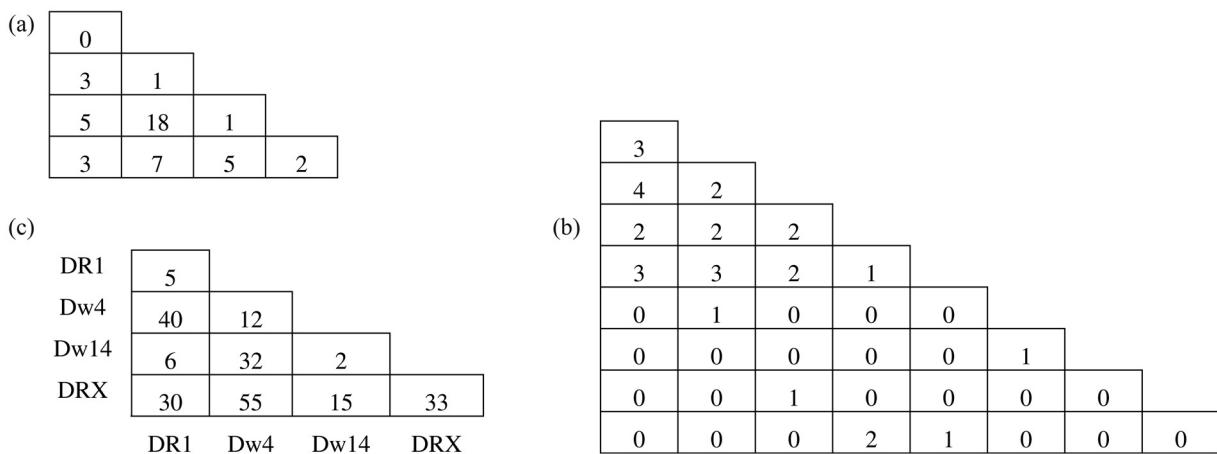


Figure 2 - Datasets for numerical experiments, given by Louis and Dempster (1987) (a), Guo and Thompson (1992) (b) and Wordsworth *et al.* (1992) (c).

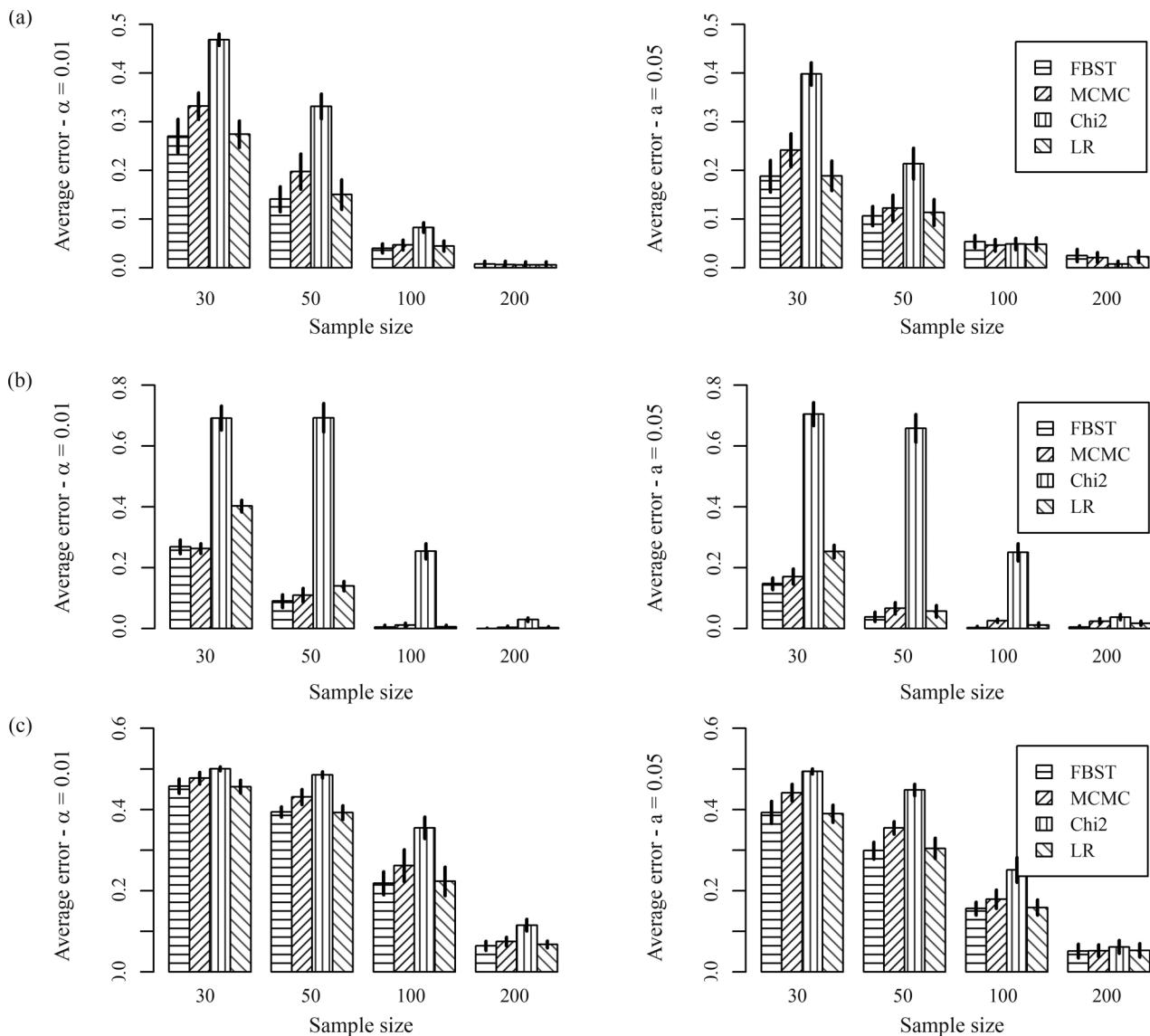


Figure 3 - Average error rates for FBST, MCMC, Chi-square and Likelihood-ratio for simulated data based on examples from Louis and Dempster (1987) (a), Guo and Thompson (1992) (b) and Wordsworth et al. (1992) (c).

are FBST and MCMC. In every case, we notice that FBST is always the best competitor (especially for small sample sizes, $n \leq 100$) or is very close to it.

These numerical results suggest that FBST is more stable than the competitors discussed in this paper, in the sense that it has good comparative performance for different datasets and allele numbers.

Final Remarks

We have introduced a simple and straightforward procedure for testing deviance from Hardy-Weinberg Equilibrium (HWE) in the presence of several alleles. This procedure was implemented in C language, and integrated into a system for parentage testing developed with FAPESP support, where it is applied in the selection of loci to be used for parentage testing. Further details of this project can

be found at <http://watson.fapesp.br/PIPEM/Pipe13/genet1.htm>. Currently, the routine is available by request to the corresponding author.

Acknowledgments

The authors are grateful for the support of EACH-USP, IQ-USP, IME-USP, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP).

References

- Borges W and Stern JM (2007) The rules of logic composition for the Bayesian epistemic e-values. Logic J IGPL 15:401-420.

- Chen JJ and Thomson G (1999) The variance for the disequilibrium coefficient in the individual Hardy-Weinberg test. *Biometrics* 55:1269-1272.
- Crow JF and Kimura M (1972) An Introduction to Population Genetics Theory. Harper & Row, New York, 591 pp.
- Crow JF (1999) Hardy, Weinberg and language impediments. *Genetics* 152:821-825.
- DeGroot MH (1970) Optimal Statistical Decisions. Wiley, New York, 489 pp.
- Emigh TH (1980) A comparison of tests for Hardy-Weinberg equilibrium. *Biometrics* 36:627-642.
- Guo SW and Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361-372.
- Hardy GH (1908) Mendelian proportions in a mixed population. *Science* 28:49-50.
- Hernández JL and Weir BS (1989) A disequilibrium coefficient approach to Hardy-Weinberg testing. *Biometrics* 45:53-70.
- Lauretto MS, Pereira CAB, Stern JM and Zacks S (2003) Full Bayesian significance test applied to multivariate normal structure models. *Braz J Probabil Stat* 7:147-168.
- Lauretto MS, Pereira CAB and Stern JM (2008) The full Bayesian significance test for mixture models: Results in gene expression clustering. *Genet Mol Res* 7:883-897.
- Levene H (1949) On a matching problem arising in genetics. *Ann Math Stat* 20:91-94.
- Louis EJ and Dempster ER (1987) An exact test for Hardy-Weinberg and multiple alleles. *Biometrics* 43:805-811.
- Madruga M, Esteves LG and Wechsler S (2001) On the Bayesianity of Pereira-Stern tests. *Test* 10:291-299.
- Montoya-Delgado LE, Irony TZ, Pereira CAB and Whittle MR (2001) An unconditional exact test for the Hardy-Weinberg equilibrium law: Sample-space ordering using the Bayes factor. *Genetics* 158:875-883.
- Pereira CAB, Nakano F, Stern JM and Whittle MR (2006) Genuine Bayesian multiallelic significance test for the Hardy-Weinberg equilibrium law. *Genet Mol Res* 5:619-631.
- Pereira CAB and Stern JM (1999) Evidence and credibility: Full Bayesian significance test for precise hypotheses. *Entropy* 1:69-80.
- Pereira CAB, Stern JM and Wechsler S (2008) Can a significance test be genuinely Bayesian? *Bayesian Anal* 3:79-100.
- Stern C (1943) The Hardy-Weinberg law. *Science* 97:137-138.
- Stern JM and Zacks S (2002) Testing the independence of Poisson variates under the Holgate bivariate distribution. The power of a new evidence test. *Stat Probabil Lett* 60:313-320.
- Stern JM (2004) Inconsistency analysis for statistical tests of hypothesis. In: López-Díaz M and Gil MA (eds) Soft Methodology and Random Information Systems. Springer, New York, pp 567-574.
- Stern JM (2007) Cognitive constructivism, eigen-solutions and sharp statistical hypotheses. *Cybernet Hum Knowing* 14:9-36.
- Weinberg W (1908) Über den Nachweis der Vererbung beim Menschen. *Jahresh Wuertt Verh Vaterl Naturkd* 64:369-382.
- Wordsworth P, Pile KD, Buckley JD, Lanchbury JSS, Ollier B, Lathrop M and Bell JI (1992) HLA heterozygosity contributes to susceptibility to rheumatoid arthritis. *Am J Hum Genet* 51:585-591.

Guest Editor: José Carlos Merino Mombach

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.